



Review

Scand J Work Environ Health [1997;23\(3\):165-178](#)

doi:10.5271/sjweh.195

Vocational outcome of intervention for low-back pain

by [van der Weide WE](#), [Verbeek JHAM](#), [van Tulder MW](#)

The following article refers to this text: [2001;27\(5\):346-352](#)

Key terms: [absenteeism](#); [evidence-based guideline](#); [guidance](#); [intervention method](#); [sick leave](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/9243726



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Vocational outcome of intervention for low-back pain

by Willeke E van der Weide, MSc,¹ Jos HAM Verbeek, PhD,¹ Maurits W van Tulder, PhD²

van der Weide WE, Verbeek JHAM, van Tulder MW. Vocational outcome of intervention for low-back pain. *Scand J Work Environ Health* 1997;23(3):165—78.

Practical management guidelines for occupational health physicians are needed for the individual support of employees with low-back pain. In this study the level of evidence regarding the efficacy of intervention with vocational outcome parameters was assessed. In a systematic literature search, 40 randomized clinical trials on different types of intervention were retrieved. Their internal validity and statistical power criteria were assessed. The randomization procedure, blinding of patients, and sample size were problematic in most studies. For patients with acute low-back pain limited or moderate evidence was found for the efficacy of no bed rest, a short period of bed rest, and spinal manipulation. For chronic patients limited evidence was found for the efficacy of antidepressants. For the other types of intervention, studies with sufficient statistical power were lacking. Such studies are needed before more-detailed evidence-based guidelines can be formulated for occupational health care.

Key terms absenteeism from work, evidence-based guidelines, guidance, intervention methods, sick leave.

One of the main tasks of the occupational health physician in The Netherlands is to offer guidance to employees who have been sick. According to the Dutch Labour Conditions Act every employer should counsel the sick employee and contract an occupational health service (in Dutch: "Arbodienst") for assistance (1). However, even for a common problem such as low-back pain, it remains unclear what counseling or intervention can prevent prolonged absenteeism from work. In order to improve the quality of occupational health care, it is important to formulate evidence-based guidelines. In 1987, the Quebec Task Force on Spinal Disorders indicated the importance of such practical management guidelines for low-back pain (2). In the United States (3), the United Kingdom (4), and The Netherlands (5), clinical guidelines have been formulated, but specific guidelines for the occupational health physician do not exist.

In the past 15 years several reviews have been published on the efficacy of various types of treatment or rehabilitation of low-back pain patients. (See, eg, references 6—10.) Most of these reviews concentrate on one form of treatment; multiple outcome variables, such as

pain, muscular strength, or functional disability, have been assessed. However, the relation of these clinical outcome variables to vocational status is unclear (11). Therefore, it is important to assess the efficacy of different types of treatment with respect to vocation separately. Vocational status parameters, such as duration of sick leave and rates of return to work, are important for the occupational health physician because employers are particularly interested in these figures.

The main objective of this study was to contribute to the development of evidence-based guidelines for occupational health physicians in their approach to the sickness absence of workers with low-back pain. Using vocational status as a measure of outcome, we reviewed the literature on nonsurgical intervention for workers with low-back pain. We have confined ourselves to the outcomes of randomized clinical trials because the randomized clinical trial is considered the most appropriate design for studying the efficacy of intervention (12). Furthermore, we evaluated studies on (sub-)acute (≤ 3 month) and chronic (>3 months) low-back pain patients separately.

¹ Coronel Institute for Occupational and Environmental Health, University of Amsterdam, Academic Medical Center, Amsterdam, The Netherlands.

² Institute for Research in Extramural Medicine, Faculty of Medicine, Vrije Universiteit, Amsterdam, The Netherlands.

Reprint requests to: Ms Willeke E van der Weide, Coronel Institute for Occupational and Environmental Health University of Amsterdam, Academic Medical Center, PO Box 22700 1100 DE Amsterdam, The Netherlands.

Methods

Retrieval of studies

A search was performed of the MEDLINE (from 1966), CLINPSYCH (from 1980) and NIOSHTIC (from 1966) data bases through December 1995. We used medical subject headings, indexing terms, and text words. The key word "back pain" was combined with one of the following terms: (employee) absenteeism, (re)employment, sick leave, return to work, sickness absence, occupational disability, and employment status. Relevant references in the retrieved articles and in published reviews were examined.

A study was included if the following criteria were met: (i) it concerned a randomized clinical trial; (ii) it concerned subjects with acute or chronic low-back pain; (iii) the intervention was directed towards acute low-back pain patients [nonsteroidal antiinflammatory drugs (NSAIDs), bed rest, spinal manipulation, back school or back exercises, and case management methods] or chronic patients (antidepressants, nonsteroidal antiinflammatory drugs, spinal manipulation, back school or exercises, behavioral therapy, and case management methods); (iv) the outcome parameter was rate of return to work, duration of sick leave or another measurement of vocational status; and (v) the article was published and available in English.

We selected these specific categories of intervention because using return to work in conjunction with all types of medical treatment did not seem feasible. Many treatments aim primarily at a reduction of pain or functional limitations. When there is no effect on functional limitations, an effect on vocational status is not expected (13). Therefore, we excluded intervention without or with evidence against their efficacy in terms of pain or functional status. Since no positive evidence was found in the literature about intervention methods for low-back pain (5, 14—21), we excluded studies on analgesics, muscle relaxants, epidural and intraarticular injections, traction, orthoses, biofeedback, acupuncture, and transcutaneous nerve stimulation. We included one exception (ie, case management methods) because of our interest in facilitating return to work. Such methods depend on the institutionalized presence of a case manager, who assesses the client's needs and ensures, through a care plan, that suitable services are provided to meet the needs, within the limits imposed by the insurance company (22). In a return-to-work process, this job is first and foremost the task of an occupational physician. We combined "back schools" and exercise therapy into one type of intervention because they overlap, since an edu-

ational program is usually combined with an exercise program.

Assessment of evidence

Reviews and meta-analyses of randomized clinical trials have offered diverse and extensive scales for assessing the quality of studies (23). However, only the randomization of treatment allocation, double-blinding, and handling of withdrawals and dropouts have proved to be associated with bias (24). Furthermore, internal validity criteria should be distinguished from other quality aspects, such as statistical power, external validity, and presentation (25). Therefore, we used the procedure represented in figure 1 in assessing the evidence and relevance for guidelines. We distinguished trials that compare a therapeutic intervention with a placebo (placebo trials) from trials that compare 2 or more different treatments (pragmatic trials). We used a broad definition for a placebo and also a categorized waiting-list group as such.

Internal validity. We assessed the internal validity of each study using the following criteria (see the appendix) (24, 26): (i) adequate randomization of allocation, (ii) blinding of study participants for treatment, (iii) comparability, or adjustment for incomparability, of the base-line characteristics³ supposedly of relevance to prognosis (27—33), (iv) loss to follow-up of less than 20%, and (v) intention-to-treat analysis.

The issue of blinding was limited to blinding of the study participants for treatment because the rate of return to work and the duration of sick leave are objective parameters. Therefore, blinding the observer seemed less relevant.

Thus the assessment of internal validity resulted in a score between 0 and 5 points. A randomized clinical trial was considered to have a high internal validity if the score was 4 or 5, and a low internal validity if it was 3 points or less. We devised detailed classification to improve the reliability of the assessment. To examine the reproducibility of the 5 items, one of us (JV) reassessed a random sample of 10 trials unaware of the assessments made by the principal assessor (WvdW). The reproducibility was sufficient (kappa 0.76, 88% agreement). The inconsistencies were mostly due to the criterion of comparability; therefore, this criterion was slightly adjusted.

Statistical power of the study. Small sample size is a major reason for erroneously concluding a finding is negative and thus missing a clinically important effect. We took this point into account with regard to studies with negative findings as a second step of the quality

³ Number of workers on sick leave, number of workers with workers' compensation, duration of complaints or sick leave, age, occupation or physical work load, radiation, previous episodes or recurrences.

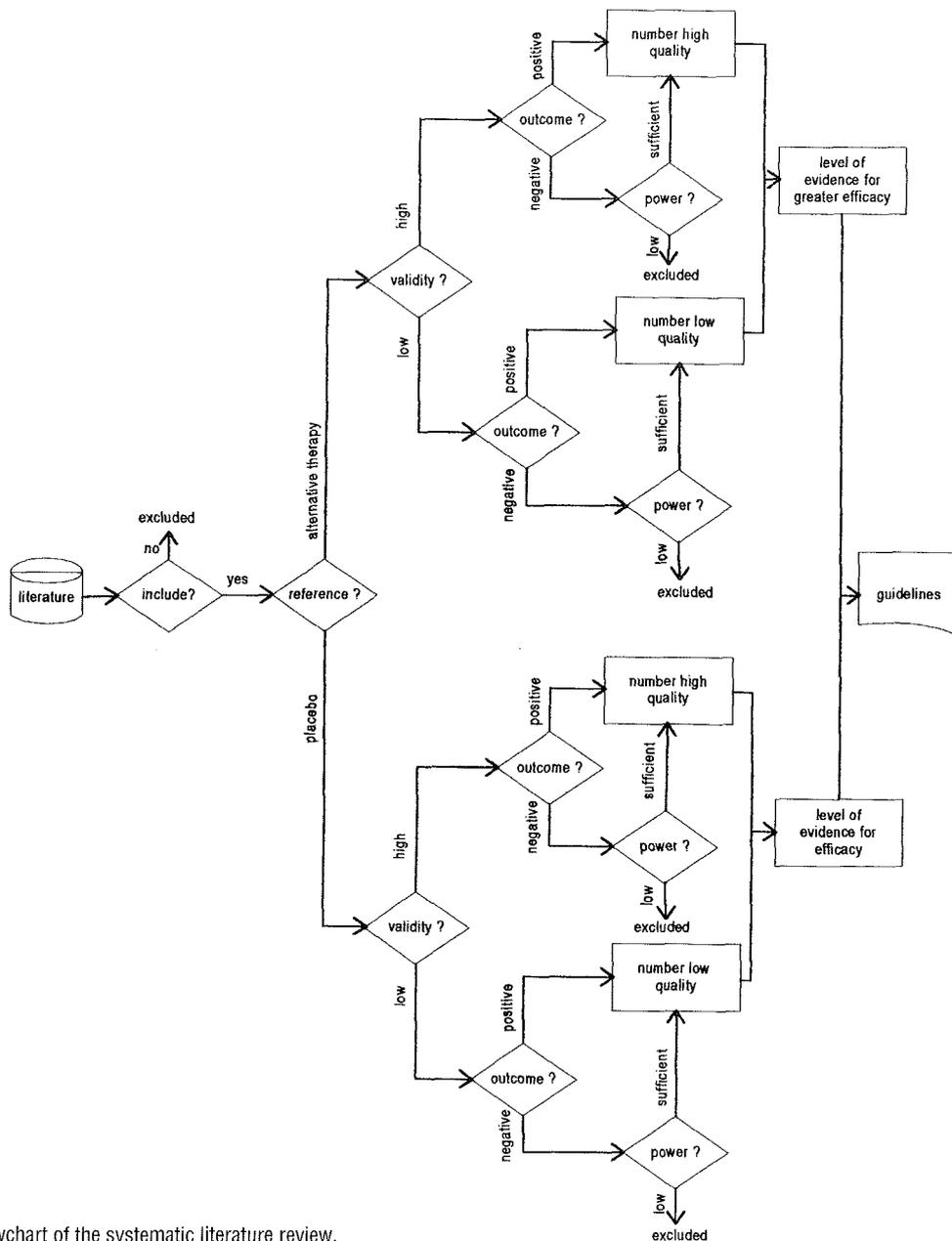


Figure 1. Flowchart of the systematic literature review.

assessment. Closely following Moher and his colleagues, we calculated the sample size necessary to detect a 25% relative difference between the treatment and reference group, with a 1-tailed $\alpha = 0.05$ and 80% power, using a z-test for proportions and a t-test for means (34). If the calculated sample size was above the real sample size, the power of the study was scored as insufficient. Studies with a negative outcome were only included in the next step of determining the level of evidence when they had enough power.

Level of evidence. For studies that used rates, we reported the rates and calculated the rate difference and rate ratio

with a 95% confidence interval (35). We reported the rates of negative events, for example, “rate of no return to work”, which means that a lower rate refers to a more effective treatment. For quantitative variables, such as sick leave, the statistical means and absolute differences between the means in the experimental and reference groups were described. We calculated confidence intervals of 95%, assuming a normal distribution.

A study can concern more than 1 outcome variable or more than 1 period in time. We presented the results for each outcome parameter. A result was defined as positive for a specific outcome parameter if there was a significant difference in favor of the experimental inter-

vention. In other words, the 95% confidence interval did not include a rate ratio of 1 or a rate difference or absolute difference of 0. A study was considered to be negative if there was no significant difference between the 2 types of intervention or if the reference intervention was more effective than the experimental intervention.

A meta-analysis is problematic with regard to the heterogeneity of the patient groups, the types of interventions, follow-up periods, or outcome parameters (36, 37). Therefore, we used a rating system with four levels of scientific evidence for an overall conclusion regarding the efficacy of the intervention. This system is an adaptation of that used in the US Clinical Practice Guideline for Acute Low-back Problems in Adults (3): level A: strong evidence: multiple high-quality randomized clinical trials; level B: moderate evidence: 1 high-quality randomized clinical trial and ≥ 1 low-quality randomized clinical trials; level C: limited evidence: 1 high-quality randomized clinical trial or multiple low-quality randomized clinical trials; level D: no evidence: 1 low-quality randomized clinical trial, no randomized clinical trial or contradictory outcomes.

Results

Of the 380 articles retrieved in our literature search, only 18 references (4%) met our inclusion criteria (38–55). Another 16 articles (56–71) were included by screening the references of the retrieved studies and of reviews about back schools (6, 7, 72), pain treatment centers (73–75), occupational rehabilitation (8, 76) and spinal manipulation (9, 10).

Some articles reported on data of the same randomized clinical trials, while others included more than 1 intervention or trial. Altogether we included 33 randomized clinical trials. In 7 of them, at least 3 different treatments were distinguished.

The number of randomized clinical trials identified for each intervention varied widely, ranging from 0 on case management methods for chronic low-back pain to 7 on back school or exercise programs for acute low-back pain.

Internal validity

Tables 1 and 2 show the internal validity scores of the studies. Only the study of Malmivaara et al (52) reached the maximum score of 5 points, while 6 randomized clinical trials had a score of 4 points. Therefore 7 randomized clinical trials had a high score for internal validity. The scores on the 5 criteria diverged widely. The randomization procedure and blinding of study participants were the most problematic to meet, being met in

only 9 (27%) and 10 (30%) randomized clinical trials, respectively. "Loss to follow-up" and "intention-to-treat" scored much better, in 31 (94%) and 28 (85%) of the trials, respectively. Adequate handling of the base-line characteristics scored 1 point in 16 randomized clinical trials (48%).

Statistical power

The outcome parameters and results are presented in tables 3 and 4. The statistical power of the studies with negative results was rated as sufficient for only 5 outcome parameters in 4 trials (15% of the negative studies). For 4 trials we estimated the power by assuming a certain — reasonable — standard deviation. In 7 trials the data were not sufficient to calculate the power for 1 or more outcome parameters.

Level of evidence in acute low-back pain studies

Nonsteroidal antiinflammatory drugs. Four studies about nonsteroidal antiinflammatory drugs were identified (58, 62, 68, 69). Three studies were placebo trials. They had a negative outcome, but none passed the statistical power criterion (58, 68, 69). (See figure 1.) Therefore no evidence for or against the efficacy of nonsteroidal antiinflammatory drugs was available for patients with acute symptoms.

Hosie's (62) pragmatic trial had only 1 point for internal validity and a negative result. We could not calculate the power of this study because only median days were reported. There was also no evidence that 1 particular nonsteroidal antiinflammatory drug was more effective than another.

Limitation of bed rest. Four studies were identified that compared different periods of bed rest or the avoidance of bed rest with a short period of bed rest (42, 52, 69, 70). Three could be used after application of the power criterion. The study with high internal validity by Malmivaara et al (52) reported positive results for duration of sick leave and negative results for the scale "ability to work". These results suggest that avoiding bed rest leads to shorter sick leave, while the difference in bed rest period did not affect functional ability. The study with low internal validity by Deyo et al (42) reported fewer days of sick leave for 2 days of bed rest in comparison with 7 days of bed rest. The other study with low internal validity, by Wiesel et al (69), reported negative results with combat trainees. In fact, the reference group with a forced period of bed rest had significantly fewer days of sick leave. A likely explanation for these contradictory results is the setting of this study. The patients, who were forced to stay in bed, were military trainees. Staying in bed until return to full duty was probably appreciated less than walking around without duty, which was permitted for the experimental group until the pain

Table 1. Internal validity of the studies on acute low-back pain.

Type of intervention	Criterion for inclusion ^a					Internal validity ^b
	Randomization procedure	Blinding of study participants	(Adjustment of in-)comparability of base-line characteristics	Loss to follow-up	Intention-to-treat analysis	
Nonsteroidal antiinflammatory drugs						
Amlie et al, 1987 (58)	0	1	1	1	0	3
Weber et al, 1993 (68)	0	1	0	1	1	3
Wiesel et al, 1980 (69)	0	0	0	1	1	2
Hosie, 1993 (62)	0	1	0	0	0	1
Short periods of bed rest						
Malmivaara et al, 1995 (52)	1	1	1	1	1	5
Wilkinson, 1995 (70)	1	0	1	1	1	4
Deyo et al, 1986 (42)	1	0	0	1	1	3
Wiesel et al, 1980 (69)	0	0	0	1	1	2
Spinal manipulation						
Blomberg et al, 1992 (41)	0	1	1	1	1	4
Wreje et al, 1992 (71)	0	1	1	1	1	4
Bergquist-Ullman & Larsson, 1977 (40)	1	0	0	1	1	3
Helliwell & Cunliffe, 1987 (61)	0	1	0	1	1	3
Waterworth & Hunter, 1985 (67)	0	0	1	1	1	3
Rasmussen, 1979 (65)	0	0	0	1	1	2
Back schools or exercise programs						
Malmivaara et al, 1995 (52)	1	1	1	1	1	5
Faas et al, 1995 (43)	1	0	1	1	1	4
Stankovic & Johnell, 1990 (53), 1995 (54)	1	0	1	1	1	4
Bergquist-Ullman & Larsson, 1997 (40)	1	0	0	1	1	3
Lindequist et al, 1984 (49)	0	0	1	1	1	3
Lindström et al, 1992 (50, 51)	0	0	1	1	1	3
Waterworth & Hunter, 1985 (67)	0	0	1	1	1	3
Case management methods						
Faas et al, 1995 (43)	1	0	1	1	1	4
Fordyce et al, 1989 (44)	0	0	0	1	1	2
Greenwood et al, 1990 (45)	1	0	0	1	0	2
Indahl, 1995 (47)	0	0	0	1	0	1

^a 0 = not adequate, 1 = adequate.^b Sum of 5 criteria.**Table 2.** Internal validity of the studies on chronic low-back pain.

Type of intervention	Criterion for inclusion ^a					Internal validity ^b
	Randomization procedure	Blinding of study participants	(Adjustment of in-)comparability of base-line characteristics	Loss to follow-up	Intention-to-treat analysis	
Antidepressants						
Alcoff et al, 1982 (57)	1	1	0	1	1	4
Nonsteroidal antiinflammatory drugs						
Meek et al, 1984 (64)	1	1	0	0	1	3
Spinal manipulation						
Arkuszewski, 1986 (59)	0	0	0	1	1	2
Gibson et al, 1985 (60)	0	1	0	1	0	2
Back schools or exercise programs						
Härkäpää et al, 1990 (46)	0	0	1	1	1	3
Keijsers et al, 1990 (48)	0	0	1	1	1	3
White, 1996 (55)	0	0	1	1	1	3
Aberg, 1984 (56)	0	0	0	1	1	2
Hurri, 1989 (63)	0	0	0	1	1	2
Behavioral therapy						
Alaranta et al, 1994 (38)	0	0	1	1	1	3
Altmaier et al, 1992 (39)	0	0	1	1	0	2
Turner, 1982 (66)	0	0	1	1	0	2

^a 0 = not adequate, 1 = adequate.^b Sum of 5 criteria.

Table 3. Results of studies on the efficacy of various conservative treatments for acute low-back pain^a. (RD = rate difference; RR = rate ratio; AD = absolute difference; 95% CI = 95% confidence interval; NS = not significant; e = experimental treatment, r = reference treatment)

Type of intervention	Treatment		Results	Con- clusion ^a	Power ^b
	Experimental	Reference			
Nonsteroidal antiinflammatory drugs					
Amlie et al, 1987 (58)	Piroxicam: 2 days 40 mg, 5 days 20 mg (N = 134)	Placebo capsules (N = 132)	Rate of no return to work after 7 days: (e) 0.69, (r) 0.79; RD: 0.10 (95% CI -0.21—0.0); RR: 0.87 (95% CI 0.75—1.01)	-	-
Weber et al, 1993 (68)	Piroxicam: 2 days 40 mg, 12 days 20 mg (N = 106)	Placebo capsules (N = 74)	Rate of no return to work at 4 weeks: (e) 0.50, (r) 0.39; RD: 0.11 (95% CI -0.04—0.26); RR: 1.28 (95% CI 0.89—1.83)	-	-
Wiesel et al, 1980 (69)	Phenylbutazone: 100 mg, 4 times/day for 5 days (e1 ^c) Aspirin: 625 mg, 4 times/day (e2 ^c)	Acetaminophen: 1 tablet, 2 times/day ^c	Mean days before return to full activity: (e1) 6.53 (SD 79) versus (r) 5.69 (SD 72) [AD: 0.84 (95% CI -1.26—2.94)] and (e2) 5.67 (SD .91) versus (r) 5.69 (SD 0.72) AD: -0.02 (95% CI -2.29—2.25)	-	-
Hosie, 1993 (62)	Topical felbinac foam and placebo capsules, 3 times/day (N = 81)	Ibuprofen 400 mg and placebo foam, 3 times/day (N = 89)	Median days taken to resume normal activities: (e) 7, (r) 6; AD: 1 (NS)	-	?
Bed rest					
Malmivaara et al, 1995 (52)	"Avoid bed rest and continue normal activities" (N = 67)	Complete bed rest of 2 days (N = 67)	Mean ability to work at 3 weeks: (e) 7.9, (r) 6.8; AD: -0.9 (95% CI -1.8—0.02) Mean ability to work at 12 weeks: (e) 8.5, (r) 7.7; AD: -0.8 (95% CI -1.5—0.1) Mean days of sick leave at 3 weeks: (e) 4.1, (r) 7.5; AD: -3.2 (95% CI -5.0—-1.3) Mean days of sick leave at 12 weeks: (e) 4.7, (r) 9.2; AD: -3.4 (95% CI -6.5—-0.2)	-	+
Wilkinson, 1995 (70)	Remain mobile, no daytime rest (N = 12)	Two days of strict bed rest (N = 10)	Mean days of sick leave at 1 month: (e) 11 (SD 10), (r) 10 (SD 8.5) AD: 1 (95% CI -6.7—8.7)	-	-
Deyo et al, 1986 (42)	Two days of bed rest recommended and usual care (N = 37)	Seven days of bed rest recommended and usual care (N = 34)	Mean days of sick leave at 3 weeks: (e) 3.1, (r) 5.6; AD: -2.5 (P = 0.01) Mean days of sick leave at 3 months: (e) 5.4, (r) 7.6; AD: -2.2 (P = 0.004)	+	+
Wiesel et al, 1980 (69)	Ambulatory, restricted duty ^d	Bed rest ^d	Mean days before return to full duty: (e) 11.8 (SD 12), (r) 6.57 (SD 23); AD: 5.2 (95% CI 4.7—5.7)	-	+
Spinal manipulation					
Blomberg et al, 1992 (41)	Described by Stoddard, changed by Evjent, Hamberg (N = 48)	Activation physical therapy drugs (N = 53)	Rate of sick leave at 1 month: (e) 0.06, (r) 0.36; RD: -0.30 (95% CI -0.44—-0.15); RR: 0.17 (95% CI 0.06—0.55) Rate of sick leave at 8 months: (e) 0.08, (r) 0.19; RD: -0.11 (95% CI -0.24—-0.03); RR: .44 (95% CI 0.15—1.32) Mean days of sick leave: (e) 25.4, (r) 58.5; AD: -33.1 (NS)	+	-
Wreje et al, 1992 (71)	Muscle energy technique and mobilization as described by Kubis (N = 18)	Placebo = massage (N = 21)	Rate of no return to work after 10 days: (e) 0.44, (r) 0.76; RD: -0.32 (95% CI -0.61—-0.02); RR: 0.58 (95% CI 0.33—1.03) Median days of sick leave at 3 weeks: (e) 7, (r) 14; AD: -7 (NS)	+	?
Bergquist-Ullman & Larsson, 1977 (40)	According to Cyriax, Kaltborn, Lewitt, Janda (N = 61)	Short-waves at lowest possible intensity (N = 66)	Rate of no return to work at 21 days: (e) 0.51, (r) 0.62; RD: -0.11 (95% CI -0.28—0.06); RR: 0.82 (95% CI 0.60—1.12) Median days of sick leave during first episode: (e) 26.5, (r) 26.5; AD: -0 (NS) Rate of recurrences during 1 year: (e) 0.51, (r) 0.54; RD: -0.03 (-0.20—0.16); RR: 0.93 (95% CI 0.67—1.30) Median days during recurrences in 1 year: (e) 7.5, (r) 8; AD: -0.5 (NS)	-	?
Helliwell & Cunliffe, 1987 (61)	Described by Cyriax (N = 6)	Placed into position for manipulation without final thrust (N = 8)	Mean days of sick leave at 1 month: (e) 6.3 (SD 5.4), (r) 6.7 (SD 5.4); AD: -0.4 (95% CI -6.1—5.3)	-	-
Waterworth & Hunter, 1985 (67)	As practiced by physical therapists (N = 24)	Nonsteroidal antiinflammatory drugs (N = 18)	Rate of no return to work within 2 weeks: (e) 0.17, (r) 0.11 RD: 0.06 (95% CI -0.15—0.26); RR: 1.50 (95% CI 0.31—7.31)	-	-
Rasmussen, 1979 (65)	Rotational (N = 12)	Short waves (N = 12)	Rate of not fit to work at 2 weeks: (e) 0.08, (r) 0.75; RD: -0.67 (95% CI -0.96—-0.38); RR: 0.11 (95% CI 0.02—0.75)	+	-
Back schools or exercise programs					
Malmivaara et al, 1995 (52)	Exercises (N = 52)	"Avoid bed rest and continue normal activities" (N = 67)	Mean ability to work at 3 weeks: (e) 7.2, (r) 7.9; AD: 0.3 (95% CI -0.5—1.2) Mean ability to work at 12 weeks: (e) 7.8, (r) 8.5; AD: 0.7 (95% CI -0.2—1.6)	-	+

continued

Table 3. Continued

Type of intervention	Treatment		Results	Con- clusion ^a	Power ^b
	Experimental	Reference			
			Mean days of sick leave at 3 weeks: (e) 5.7, (r) 4.1; AD: 1.8 (95% CI 0.1—3.5)	-	-
			Mean days of sick leave at 12 weeks: (e) 7.2, (r) 4.7; AD: 2.5 (95% CI 0.2—4.9)	-	-
Faas et al, 1995 (43)	Exercise therapy (N = 122)	Ultrasound at lowest possible dose (N = 150)	Rate of at least one back-related day lost during back pain in 1 year: (e) 0.84, (r) 0.77; RD: 0.07 (95% CI -0.03—0.17); RR: 1.09 (95% CI 0.96—1.24)	-	+
			Mean days of sick leave during 1 year: (e) 22 (SD 41), (r) 20 (SD 34); AD: 2 (95% CI -8—12)	-	-
Stankovic & Johnell, 1990, 1995 (53, 54)	"Mini" back school (N = 50)	McKenzie method (N = 50)	Rate of no return to work during initial 4 weeks: (e) 0.20, (r) 0.02; RD: 0.18 (95% CI 0.06—0.30); RR: 10.00 (1.33—75.23)	-	-
			Mean days of sick leave during initial period: (e) 21.6 (SD 15.3), (r) 11.9 (SD 6.5); AD: 9.7 (95% CI -5.1—14.3)	-	-
			Mean days of sick leave during recurrences in 1 year: (e) 40.1 (SD 59.6), (r) 27.0 (SD 14.4); AD: 13 (95% CI -3.9—30)	-	-
			Mean days of sick leave during following 4 years: (e) 103.9 (SD 181.5), (r) 84.1 (SD 144.3); AD: 19.8 (95% CI -49—88)	-	-
Bergquist-Ullman & Larsson, 1977 (40)	Back school (N = 55)	Short-waves at lowest possible intensity (N = 66)	Rate of no return to work at 21 days: (e) 0.33, (r) 0.62; RD: -0.29 (95% CI -0.46—-0.12); RR: 0.53 (95% CI 0.35—0.80)	+	-
			Median days of sick leave during first episode: (e) 20.5, (r) 26.5; AD: -6 (NS)	-	?
			Rate of recurrences during 1 year: (e) 0.49, (r) 0.54; RD: -0.05 (-0.23—0.12); RR: 0.90 (95% CI 0.64—1.27)	-	-
			Median days during recurrences in 1 year: (e) 7.5, (r) 8; AD: -0.5 (NS)	-	?
Lindequist et al, 1984 (49)	Back school (N = 24)	Less intense attention (N = 32)	Mean days of sick leave in first episode: (e) 31, (r) 29; AD: 2 (NS)	-	(-)
			Mean days of sick leave during recurrences during 1 year: (e) 36, (r) 39; AD: -3 (NS)	-	(-)
Lindström et al, 1992, 1992 (50, 51)	Graded activity program (N = 51)	Usual care (N = 52)	Mean days of sick leave for low-back pain in first episode: (e) 70 (SD 89), (r) 106 (SD 109), AD: -36 (95% CI -74—2.4)	-	-
			Mean days of sick leave for low-back pain during 2 years: (e) 85 (SD 129), (r) 137 (SD 145); AD: -52 (95% CI -105—0)	+	-
Waterworth & Hunter, 1985 (67)	Exercises (N = 16)	Nonsteroidal antiinflammatory drugs (N = 18)	Rate of no return to work within 2 weeks: (e) 0.25, (r) 0.11; RD: 0.14 (95% CI -0.12—0.40); RR: 2.25 (95% CI 0.47—10.69)	-	-
Case management methods					
Faas et al, 1995 (43)	Care by general practitioner, trained to discuss matters of low-back pain and analgesics (N = 122)	Ultrasound at lowest possible dose (N = 150)	Rate of at least one back-related day lost during back pain in 1 year: (e) 0.75, (r) 0.77; RD: -0.02 (95% CI -0.13—0.09); RR: 0.97 (95% CI 0.84—1.13)	-	+
			Mean days of sick leave during 1 year: (e) 23 (SD 55), (r) 20 (SD 34); AD: 3 (95% CI -9—15)	-	-
Fordyce et al, 1986 (44)	Behavioral management, time-based (N = 57)	Traditional pain-based management (N = 50)	Mean vocational status at 6 weeks: NS (no further report)	-	?
			Mean vocational status at 9—12 months: (e) 0.47 (SD 1.25), (r) 0.87 (SD 1.77); AD: -0.40 (95% CI -0.99—0.19)	-	-
Greenwood et al, 1990 (45)	Health and psychosocial case-management approach in second week (N = 121)	Usual care (N = 163)	Mean days of sick leave until return to work: (e) 102.6 (SD 147.9), (r) 109.7 (SD 148.7); AD: 7.1 (95% CI -28—42)	-	-
Indahl, 1995 (47)	Treating low-back pain as benign, self-limiting condition (N = 463)	Conventional medical system (N = 512)	Rate of no return to work at 200 days: (e) 0.30, (r) 0.60; RD: -0.30 (95% CI -0.36—-0.24); RR: 0.50 (95% CI 0.43—0.59)	+	-
			Proportional hazard rate = 2.23, P = 0.000	+	-

^a + = positive conclusion (experimental treatment better than reference treatment and 95% CI of RD does not include 0 or 95% CI of RR does not include 1 sufficient power: sample size enough to detect a 25% difference with 1-tail $\alpha = 0.05$ and $1-\beta = 0.80$), - = negative conclusion.

^b + = sufficient power, - = insufficient power, () = estimated power, ? = power could not be established.

^c Experimental + reference groups = 45.

^d Experimental + reference groups = 80.

Table 4. Results of studies on the efficacy of various conservative treatments for chronic low-back pain. (RD = rate difference; RR = rate ratio; AD = absolute difference; 95% CI = 95% confidence interval; NS = not significant; e = experimental treatment; r = reference treatment; NS = not significant)

Type of intervention	Treatment		Results	Con- clusion ^a	Power ^b
	Experimental	Reference			
Antidepressants					
Alcott et al, 1982 (57)	Imipramine (Tofranil) 75 mg, 1 pill a night, after 3 days 2 pills a night (N = 28) for 8 weeks	Placebo capsules (N = 22)	Mean limitation of work at 8 weeks: AD: ? (P = 0.004)	+	
Nonsteroidal antiinflammatory drugs					
Meek et al, 1984 (64)	Colchicine, 1 mg intravenous on day 1 and 0.6 mg capsules for 14 days ^c	Placebo, intravenous and capsules ^c	Mean severity index of work impairment: P < 0.01	+	
Spinal manipulation					
Arkuszewski, 1986 (59)	Traction and manipulation, based on Lewit (N = 47)	Bed rest, analgesics, massage (N = 41)	Rate of no return to previous job at 6 months: (e) 0.40, (r) 0.64; RD: -0.24 (95% CI -0.43—-0.03); RR: 0.64 (95% CI 0.42-0.97)	+	
Gibson et al, 1985 (60)	Osteopathic treatment (N = 41)	Detuned short-wave diathermy (N = 34)	Prepost differences in rates unable to work or with restricted activities: at 2 weeks: (e) -0.02, (r) -0.12, prepost difference: 0.10; at 4 weeks: (e) -0.05, (r) -0.20, prepost difference: 0.15; at 12 weeks: (e) -0.10, (r) -0.31, prepost difference: 0.21	-	-
Back schools or exercise programs					
Härkäpää et al, 1990 (46)	Three weeks inpatient, modified Swedish back school, refresher course after 1½ years (N = 157) (e1) Modified Swedish back school, refresher course after 1½ years (N = 159) (e2)	Written and oral instructions (N = 160) (r)	e1 versus r: mean scores of differences of sickness allowance days between pretreatment and 1.5 years posttreatment due to spinal disorders: (e) +0.3, (r) +3.8, AD: -3.5 (NS) and 2.5 years posttreatment due to spinal disorders: (e) +1.6, (r) +4.8, AD: -3.2 (NS) e2 versus r: mean scores of differences of sickness allowance days between pretreatment and 1.5 years posttreatment due to spinal disorders: (e) -0.2, (r) +3.8, AD: -4.0 (NS) and 2.5 years posttreatment due to spinal disorders: (e) +2.2, (r) +4.8, AD: -2.6 (NS) e1 versus e2: mean scores of differences of sickness allowance days between pretreatment and 1.5 years posttreatment due to spinal disorders: AD: +0.5 (NS) and 2.5 years posttreatment due to spinal disorders: AD: -0.6 (NS) Rate of disability pension after 4.5 years on the average: e1 versus r: (e) 0.10, (r) 0.12, RD: -0.02 (-0.09—0.05); RR: 0.86 (95% CI 0.46—1.61); e2 versus r: (e) 0.08, (r) 0.12, RD: -0.04 (95% CI -0.10—0.03); RR: 0.69 (95% CI 0.35—1.35); e1 versus e2: RD: 0.02 (95% CI -0.04—0.08); RR: 1.25 (95% CI 0.62—2.50)	-	(-)
Keijsers et al, 1990 (48)	Maastricht back school ^d	Waiting list ^d	Mean days of sick leave at 6 months: (e) 19.0 ± 43.2, (r) 3.43 ± 10.2; AD: 15.6 (95% CI 1.6—29.5)	-	-
White, 1966 (55)	Graded activity program (N = 99)	Usual care (N = 95)	Rate of satisfactory results at 3 months (ability to work full time, to work at preaccident job with up to 20% time loss, or to work at modified work): (e) 0.42, (r) 0.16; RD: -0.27 (95% CI -0.39—-0.14); RR: 0.68 (95% CI 0.57—0.83)	+	
Aberg, 1984 (56)	Inpatient rehabilitation program (N = 164)	No specification (N = 189)	Rate of no change in vocational factors at 4 months: (e) 0.85, (r) 0.91; RD: -0.06 (95% CI -1.3—0.01); RR: 0.93 (95% CI 0.87—1.00) Rate of no change in vocational factors at 8 months: NS (no further report)	-	+
Hurri, 1989 (63)	Swedish back-school (N = 95)	Written instructions and usual care (N = 93)	Mean days of sick leave: AD: ? (NS); RD: ? Rate with various durations of sick leave: RD: ? (NS); RR: ?	-	?
Behavioral therapy					
Alaranta et al, 1994 (38)	Three weeks home-training and 3 weeks inpatient physical and psychological program (N = 152)	Three-week inpatient program light physical exercises (N = 141)	Mean days of sick leave during 1 year: (e) 33.9, (r) 36.9; AD: -3 (NS) Rate of WHO occupational handicap classification 3—5 versus 0—2 at 1 year: (e) 0.10, (r) 0.16; RD: -0.06 (95% CI -0.14—0.02); RR: 0.63 (95% CI 0.34—1.17)	-	(-)

continued

Table 4. Continued

Type of intervention	Treatment		Results	Con- clusion ^a	Power ^b
	Experimental	Reference			
Altmaier et al, 1992 (39)	Three weeks inpatient rehabilitation program with psychological intervention (N = 24)	Three-week inpatient rehabilitation program without psychological intervention (N = 21)	Rate of no return to work (full-time) at 6 months: (e) 0.52, (r) 0.33; RD: 0.19 (95% CI -0.10—-0.48); RR: 1.57 (95% CI 0.76—3.26) Rate of no return to work (full-time or part-time) at 6 months: (e) 0.24, (r) 0.14; RD: 0.10 (95% CI -0.14—0.33); RR: 1.67 (95% CI 0.46—6.10)	-	-
Turner, 1982 (66)	Progressive relaxation training (N = 14) (e1) Cognitive-behavioral group therapy (N = 13) (e2)	Waiting list (N = 9)	Mean hours worked per week posttreatment: (e1): (e) 21.9 (SD 17.9), (r) 22.0 (SD 16.7); AD: 0.1 (95% CI -14.3—14.5); (e2): (e) 14.8 (SD 14.0), (r) 22.0 (SD 16.7); AD: 7.2 (95% CI -6.1—20.5); (e1) versus (e2): AD: -3.4 (-15.4—8.6) Mean hours worked per week at 1 ½ to 2 years for (e1) versus (e2): (e1) 22.8 (SD 18.9), (e2) 38.0 (SD 17.7); AD: 15.2 (0.54—29.9)	- - - + (e2)	- - - -

^a + = positive conclusion (experimental treatment better than reference treatment and 95% CI of RD does not include 0 or 95% CI of RR does not include 1 sufficient power: sample size enough to detect a 25% difference with 1-tail $\alpha = 0.05$ and $1-\beta = 0.80$), - = negative conclusion.

^b + = sufficient power, - = insufficient power, () estimated power, ? = power could not be established.

^c Experimental + reference groups = 25.

^d Experimental + reference groups = 77.

disappeared. Therefore, for "civilian life" there is moderate evidence for the efficacy of avoiding bed rest or for short periods of bed rest, in terms of duration of sick leave after 3 months for patients with and without radiating pain.

Spinal manipulation. Six randomized clinical trials about spinal manipulation were identified (40, 41, 61, 65, 67, 71). In three a placebo group was used. One study qualified for the final assessment of the level of evidence. It had a high internal validity score and reported positive results after 10 days for patients with pelvic joint dysfunction (71).

Two of the 3 pragmatic trials reached the final assessment. The trial by Blomberg et al (41) had high internal validity and a positive result for patients with or without radiating pain. In a study with low internal validity Rasmussen (65) reported that spinal manipulation was more effective than short waves for patients without signs of root pressure (65).

In conclusion, there is limited evidence for the efficacy of spinal manipulation in comparison with placebo treatment in cases of pelvic joint dysfunction. There is moderate evidence that spinal manipulation is more effective in the short run than other conservative types of treatment, like physiotherapy, at least for patients without radiating pain.

Back schools or exercise therapy. Seven randomized clinical trials were identified in 9 articles (40, 43, 49—54, 67). Both placebo trials could be used for the assessment of the level of evidence. One randomized clinical trial had high internal validity with a negative result (43) and the other had low internal validity with a positive result (40). The differences between these 2 studies could not be explained by the intensity of the program. While the

program of the negative trial had 10 sessions, the positive one had only 4. Merely the content of the program differed, the program of the positive study including a worksite visit.

Five randomized clinical trials compared 2 or more alternative interventions, and 2 of these trials could be used for the level of evidence. One study with high internal validity had a negative result after 3 and 12 weeks of follow-up (52). The study by Lindström et al (50, 51) had low internal validity and reported a positive result with 2 years of follow-up. The intervention of these trials differed strongly in intensity in that the negative study had an exercise program of 1 session and the positive trial comprised a program of 10 sessions, on the average, and included a worksite visit.

In conclusion, there is no evidence for the efficacy of back schools or exercise therapy, and there is no evidence either that back schools or exercise therapy is more effective than usual medical care.

Case management methods. We identified 4 randomized clinical trials using case management methods (43—45, 47). The only placebo trial, a high-quality study, reported a negative result (43). Only 1 of the 3 pragmatic trials, a study with low internal validity, could be used in the final assessment of the level of evidence (47). It reported a positive outcome. In conclusion, there was neither evidence for the efficacy of case management methods nor for the greater efficacy of case management in comparison with conventional treatment.

Level of evidence in chronic low-back pain studies

Antidepressants. Only one placebo randomized clinical trial (57), a study with high internal validity and a positive result, was identified that used antidepressants. Therefore there was limited evidence of efficacy for

chronic low-back pain patients after 2 months of antidepressants.

Nonsteroidal antiinflammatory drugs. Only 1 placebo trial with low internal validity and with a positive result was identified (64), therefore there was no evidence for or against the effectiveness of nonsteroidal antiinflammatory drugs for chronic patients.

Spinal manipulation. Two articles about spinal manipulation were identified (59, 60). One study used placebo treatment, but had insufficient statistical power (60). The other trial compared alternative treatments (59). It was a study with low validity and a positive outcome.

In conclusion, there is no evidence for the efficacy of spinal manipulation for chronic patients, whether compared with a placebo or other treatments.

Back schools or exercise therapy. Five randomized clinical trials were identified (46, 48, 55, 56, 63) for back schools or exercise therapy. One of these trials used a waiting list group, but did not pass the statistical power criterion.

One study had no description of the treatment for the reference group (56). This study was included in the group of pragmatic trials, with the other three. Two trials, both with low internal validity, were considered for the final assessment. One reported positive results (55), the other negative (56). Both were inpatient programs in a rehabilitation center and lasted 6 weeks. These contradictory results indicate that there is no evidence that back schools or exercise therapy is more effective than usual care.

Behavioral therapy. Three randomized clinical trials with low internal validity were identified that used behavioral therapy (38, 39, 66). In 1 randomized clinical trial, Turner (66) compared 2 groups subjected to different behavioral approaches with a group on a waiting list, but the 2 behavioral approach groups were also compared with each other. Only the last comparison could be used for the final assessment of the level of evidence. On the basis of this study it was concluded that there is no evidence that cognitive-behavioral group therapy as part of progressive-relaxation training is more effective in the long term than progressive-relaxation training alone.

Case management methods. No randomized clinical trials about case management were identified for chronic low-back pain patients.

Discussion

We used 2 strategies for identifying relevant articles for our review. First, we studied other reviews on interven-

tion for low-back pain patients, and, second, we searched computer data bases. An intensive search is important because evidence for guidelines is based on only 1 or 2 articles. Therefore, a conclusion can easily be reversed as 1 or 2 new articles on an intervention program are identified. Our key words were broad, but failed to select all articles with vocational status outcome parameters from MEDLINE and the other data bases. A data-base search with key words regarding specific types of intervention would be more useful. However, a key word search can also not guarantee the identification of all relevant articles. Publication bias, the limited number of indexed journals, and the use of different key words are all problems in a literature review that should be reduced, for instance, by a registration system of clinical trials (77). To this end, the Cochrane Collaboration (25), which aims at the registration of all randomized clinical trials, is promising.

We used a stepwise quality assessment, starting with an assessment of the internal validity. Ideally, studies with low internal validity should not be included because of the association of methodological inconsistencies with bias (24). An obvious disadvantage is the lack of studies with high internal validity, as was concluded in other reviews (20). If the results of studies support each other and there is a lack of studies with high internal validity, trials with low internal validity are the second best option for establishing the level of evidence. The use of the 5 internal validity criteria is comparable to the use of the criteria that are part of the quality assessment by other authors (23). The condition that 4 or 5 criteria should be met for high internal validity is arbitrary. We argue that at least 4 criteria should be met because of the proved association between the selected criteria and bias (24). However, the subjectivity of quality assessments remains a problem. Therefore, the development of an easy, universal, and reliable scale deserves attention (23). For instance, we did not weight the criteria because the relative value of each criterion was arbitrary. The choice of the rating system for the level of evidence was also subjective. The choice of our system corresponds as much as possible with the AHCPR system of the Agency for Health Care Policy and Research (AHCPR) in the United States although it could not be copied completely (3). We used randomized clinical trials and no other studies or expert opinions in the same manner as the AHCPR system. It should be noted that a more conservative system, such as that used by Tulder et al (20, 21), would lead to less evidence and an even greater demand for more randomized clinical trials. Therefore, we would encourage a discussion about the balance between a demand for more research and satisfactory evidence for practical guidelines.

The methodological problems most prevalent in the studies concerned adequate randomization and blind-

ing of study participants. This result is consistent with those of reviews on other outcome parameters, such as pain and functional limitations (20). However, "intention to treat" scored much better among the retrieved studies in our review. The most likely explanation is the use of sick leave rates as an outcome parameter, which could be retrieved despite the failure to follow-up the respondents.

The second step in the quality assessment was related to the number of study participants. In meta-analyses, multiple studies with low power can be combined to show one, more powerful outcome. However, heterogeneity as to the intervention, follow-up period, study population, and the like does not always permit such a combination (36, 37). In our view, the reviewed intervention studies differ too much for a meta-analysis, and we therefore looked at the studies separately. We excluded studies with low power to prevent an erroneous negative conclusion about the efficacy. Unfortunately, many identified studies were excluded because of insufficient power, despite the fact that we used a 25% difference between the experimental and reference treatment as a relevant clinical difference. In our opinion, this criterion is not very rigorous. For instance, if we had considered a smaller difference as relevant (eg, 20%) even more studies would have been excluded. Therefore, we recommend sample size calculations at the beginning of a study to minimize this problem as much as possible. These calculations give only a global estimation of the sample size because of the uncertainty about some parameters. The lack of power for the vocational outcome parameter in the retrieved studies can be due to the fact that some studies were primarily aimed at other outcome parameters, with less variance.

Most of the studies have broad confidence intervals for their outcome parameters. This great variance can be partially explained by the small number of patients included, but also by the assessment of differences in the means. Although we assumed a normal distribution for duration of sick leave, as most of the studies did, it is more appropriate to use survival techniques for analyzing time to return to work. Then fewer patients would be necessary to achieve sufficient power.

We have presented the outcome of the trials in 2 ways, as an absolute outcome (rate differences for rates) and a relative outcome (rate ratios for rates). For some studies it was not possible to calculate confidence intervals, sometimes not even the outcome itself. In these articles, the authors stated that the differences were not significant, without reporting standard deviations. In our opinion, it is important to report results uniformly to allow for comparison of the outcomes. As argued before, we recommend a standard method of reporting results for time to return to work.

Another explanation for the broad confidence intervals is the fact that return to work is a multifactorial process, with many confounders. Its outcome is not only influenced by the patient and his or her low-back pain, but also by the work situation, the possibilities for working part-time or working at a lower pace, and problems in the home situation and even the macro-economic situation (78, 79). Intervention that is directed at the patient cannot influence all these factors; it should therefore have a high impact on the patient to have a net positive outcome. Because of this multifactorial influence on sick leave, a multifactorial intervention strategy can be the most effective. The individual patient with low-back pain should be advised of effective treatment options. However, at the same time, the work situation should be adapted to make it easier for the patient to return to work. A good example of this "case management approach" is the Sherbrooke model, which includes both clinical and ergonomic approaches (80). More randomized clinical trials on this subject are needed, as has already been noted by Lechner et al (46) in their review about the effectiveness of work hardening and work conditioning programs (76). They found only 2 randomized clinical trials out of 12, most of them with promising results.

Recently, another relevant review was published by Scheer et al (81) about randomized clinical trials for acute low-back pain in relation to return to work. They assessed the methodological quality of the studies, but the relationship between this quality rating and the level of evidence remained unclear. Furthermore, we found more randomized clinical trials on spinal manipulation which led to evidence for the efficacy of this intervention. For bed rest, exercise, back schools and case management methods, their conclusion was the same as ours. However, with our rating system, we showed that one of the main reasons for negative findings is lack of statistical power.

Concluding remarks

For most of the reviewed types of intervention, the scientific evidence is limited for the efficacy of intervention for patients with low-back pain in terms of sickness absence rates or duration of sick leave. We have formulated the following main points from this review for the guidelines for occupational physicians:

- bed rest should be limited or even avoided; normal activity should be continued as much as possible;
- if any conservative treatment for patients with acute low-back pain is considered, spinal manipulation is the best option;
- antidepressants can be helpful for chronic low-back pain patients.

Furthermore, there are promising results for exercise and education programs, especially for intensive programs in an occupational setting (6). However, these positive findings have not yet been confirmed in randomized clinical trials with vocational outcome parameters. More high-quality randomized clinical trials with sufficient statistical power and vocational outcome parameters are still needed before guidelines based on stronger evidence can be established for occupational physicians.

Acknowledgments

We thank Dr GJ Bonsel from the Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, for his valuable comments on the manuscript.

References

1. Poest PE van der, Boere AHM, editors. *Handboek Arboret: rechten en verplichtingen toegelicht voor de werkgever en werknemer*. Den Haag: Sdu uitgeverij Plantijnstraat, 1994.
2. Spitzer WO, LeBlanc FE, Dupuis M, editors. *Scientific approach to the assessment and management of activity-related spinal disorders: a monograph for clinicians: report of the Quebec Task Force on Spinal Disorders*. *Spine* 1987;12 Suppl: 1—59.
3. Bigos S, Bowyer O, Braen G, Brown KC, Deyo RA, Halde- man S, et al. *Acute low back problems in adults: clinical practice guideline, quick reference guide number 14*. Rockville (MD): US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, 1994. AHCPR Pub No 95—0643.
4. Rosen M, Breen A, Hamann W, Harker P, Jayson MIV, Kelly E, et al. *Report of a clinical standards advisory group committee on back pain*. London: Her Majesty's Stationery Office, 1994.
5. Faas A, Chavannes AW, Koes BW, Hoogen JMM van den, Mens JMA, Smeele LJM, et al. *Practice guideline 'low back pain'*. Utrecht: Dutch College of General Practitioners, 1996.
6. Koes BW, Tulder MW van, Windt DAWM van der, Bouter LM. The efficacy of back schools: a review of randomized clinical trials. *J Clin Epidemiol* 1994;47:851—62.
7. Linton SJ, Kamwendo K. Low back schools: a critical review. *Phys Ther* 1987;67:1375—83.
8. Feuerstein M, Menz L, Zastowny T, Barron BA. Chronic back pain and work disability: vocational outcomes following multi-disciplinary rehabilitation. *J Occup Rehabil* 1994;4:229—51.
9. Koes BW, Assendelft WJJ, Heijden GJMG van der, Bouter LM, Knipschild PG. Spinal manipulation and mobilization for back and neck pain: a blinded review. *BMJ* 1991;303:1298—303.
10. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Brook RH. Spinal manipulation for low-back pain [review]. *Ann Intern Med* 1992;117:590—8.
11. Hazard RG, Haugh LD, Green PA, Jones PL. Chronic low back pain: the relationship between patient satisfaction and pain, impairment, and disability outcomes. *Spine* 1994;19: 881—7.
12. Cook DJ, Guyatt GH, Laupacis A, Sackett DI. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992;102 suppl:305—11.
13. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med* 1994;38:1—4.
14. Frymoyer JW, editor. *The adult spine: principles and practice*. New York (NY): Raven Press Ltd, 1991.
15. Jayson MIV, editor. *The lumbar spine and back pain*. New York (NY): Churchill Livingstone, 1987.
16. Deyo RA. Non-operative treatment of low back disorders: differentiating useful from useless therapy. In: Frymoyer JW, editor. *The adult spine: principles and practice*. New York (NY): Raven Press Ltd, 1991:1567—80.
17. Deyo RA. Conservative therapy for low back pain. *JAMA* 1983;250:1057—62.
18. Koes BW, Hoogen HMM van den. Efficacy of bed rest and orthoses of low-back pain: a review of randomized clinical trials. *Eur J Phys Med Rehabil* 1994;4:86—93.
19. Pope MH. Orthoses for treatment of low back disorders. In: Frymoyer JW, editor. *The adult spine: principles and practice*. New York (NY): Raven Press Ltd, 1991:1607—15.
20. Tulder MW van, Koes BW, Bouter LM. Conservative treatment of chronic low back pain: a systematic mega-review of 80 randomized clinical trials of 14 interventions. In: Tulder MW van, Koes BW, Bouter LM, editors. *Low back pain in primary care*. Amsterdam: EMGO Institute, 1996:245—85.
21. Tulder MW van, Koes BW, Bouter LM. Conservative treatment of acute low back pain: a systematic mega-review of 81 randomized clinical trials of 11 interventions. In: Tulder MW van, Koes BW, Bouter LM, editors. *Low back pain in primary care*. Amsterdam: EMGO Institute, 1996:205—43.
22. Shepherd G. Case management. *Health Trends* 1990;22:59—61.
23. Moher D, Jadad A, Nichol G, Penman M, Tugwell T, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clin Trials* 1995;16:62—73.
24. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408—12.
25. Oxman AD, editor. Section VI: preparing and maintaining systematic reviews: Cochrane collaboration handbook. Oxford: Cochrane Collaboration, 1994.
26. Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports on randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;17:1—12.
27. Abenhaim L, Suissa S, Rossignol M. Risk of recurrence of occupational back pain over three year follow up. *Br J Ind Med* 1988;45:829—33.
28. Coste J, Delecoeuillerie G, Cohen de Lara A, Le Parc JM, Paolaggi JB. Clinical course and prognostic factors in acute low back pain: an inception cohort study in primary care practice. *BMJ* 1994;308:577—80.
29. Fredrickson BE, Trief PM, VanBeveren P, Yuan HA, Baum G. Rehabilitation of the patient with chronic back pain: a search for outcome predictors. *Spine* 1988;13:351—3.
30. Goertz MN. Prognostic indicators for acute low-back pain. *Spine* 1990;15:1307—10.
31. Hellsing AL, Linton SJ, Källemark M. A prospective study of patients with acute back and neck pain in Sweden. *Phys Ther* 1994;74:116—24.

32. Lancourt J, Kettelhut M. Predicting return to work for lower back pain patients receiving worker's compensation. *Spine* 1992;17:629—38.
33. Lehmann TR, Spratt KF, Lehmann KK. Predicting long-term disability in low back injured workers presenting to a spine consultant. *Spine* 1993;18:1103—12.
34. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122—4.
35. Rothman KJ. *Modern epidemiology*. Boston/Toronto: Little, Brown and Company, 1986:164—75.
36. Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E, Canadian Task Force on the Periodic Health Examination. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. *J Clin Epidemiol* 1990;43:891—905.
37. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;309:1351—5.
38. Alaranta H, Rytökoski U, Rissanen A, Talo S, Rönnemaa T, Puukka P, et al. Intensive physical and psychosocial training program for patients with chronic low back pain: a controlled clinical trial. *Spine* 1994;19:1339—49.
39. Altmaier EM, Lehmann TR, Russell DW, Weinstein JN, Kao CF. The effectiveness of psychological interventions for the rehabilitation of low back pain: a randomized controlled trial evaluation. *Pain* 1992;49:329—35.
40. Bergquist-Ullman M, Larsson U. Acute low back pain in industry: a controlled prospective study with special reference to therapy and confounding factors. *Acta Orthop Scand* 1977;170:1—117.
41. Blomberg S, Svärdsudd K, Mildenerger F. A controlled, multicentre trial of manual therapy in low-back pain: initial status, sick-leave and pain score during follow-up. *Scand J Prim Health Care* 1992;10:170—8.
42. Deyo RA, Diehl AK, Rosenthal M. How many days of bed rest for acute low back pain?: a randomized clinical trial. *N Engl J Med* 1986;315:1064—70.
43. Faas A, van Eijk JTM, Chavannes AW, Gubbels JW. A randomized trial of exercise therapy in patients with acute low back pain: efficacy sickness absence. *Spine* 1995;20: 941—7.
44. Fordyce WE, Brockway JA, Bergman JA, Spengler D. Acute back pain: a control-group comparison of behavioral vs traditional management methods. *J Behav Med* 1986;9:127—40.
45. Greenwood JG, Wolf HJ, Pearson RJ, Woon CL, Posey P, Main CF. Early intervention in low back disability among coal miners in West Virginia: negative findings. *J Occup Med* 1990;32:1047—52.
46. Härköpää K, Mellin G, Järvikoski A, Hurri H. A controlled study on the outcome of inpatient and outpatient treatment of low back pain: part III: long-term follow-up of pain, disability, and compliance. *Scand J Rehabil Med* 1990;22:181—8.
47. Indahl A, Velund L, Reikeraas O. Good prognosis for low back pain when left untampered: a randomized clinical trial. *Spine* 1995;20:473—7.
48. Keijsers JFME, Steenbakkens WHL, Meertens RM, Bouter LM, Kok GJ. The efficacy of the back school: a randomized trial. *Arthritis Care Res* 1990;3:204—9.
49. Lindequist S, Lundberg B, Wikmark R, Bergstad B, Lööf G, Ottermark AC. Information and regime at low back pain. *Scand J Rehabil Med* 1984;16:113—6.
50. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. *Phys Ther* 1992;72:279—90.
51. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Nachemson A. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain: a randomized prospective clinical study with a behavioral therapy approach. *Spine* 1992;17:641—52.
52. Malmivaara A, Häkkinen U, Aro T, Heinrichs M-L, Koskeniemi L, Kuosma E, et al. The treatment of acute low back pain — bed rest, exercises, or ordinary activity? *N Engl J Med* 1995;332:351—5.
53. Stankovic R, Johnell O. Conservative treatment of acute low-back pain: a prospective randomized trial: McKenzie method of treatment versus patient education in "mini back school". *Spine* 1990;15:120—3.
54. Stankovic R, Johnell O. Conservative treatment of acute low-back pain: a 5-year follow-up study of two methods of treatment. *Spine* 1995;20:469—72.
55. White AWM. Low back pain in men receiving workmen's compensation. *Can Med Assoc J* 1966;95:50—6.
56. Aberg J. Evaluation of an advanced back pain rehabilitation program. *Spine* 1984;9:317—8.
57. Alcock J, Jones E, Rust P, Newman R. Controlled trial of imipramine for chronic low back pain. *J Fam Pract* 1982;14: 841—6.
58. Amlie E, Weber H, Holme I. Treatment of acute low-back pain with piroxicam: results of a double-blind placebo-controlled trial. *Spine* 1987;12:473—6.
59. Arkuszewski Z. The efficacy of manual treatment in low back pain: a clinical trial. *Man Med* 1986;2:68—71.
60. Gibson T, Grahame R, Harkness J, Woo P, Blagrove P, Hills R. Controlled comparison of short-wave diathermy treatment with osteopathic treatment in non-specific low-back pain. *Lancet* 1985;1258—61.
61. Helliwell PS, Cunliffe G. Manipulation in low back pain. *Physician* 1987;April:187—8.
62. Hosie GAC. The topical NSAID, felbinac, versus oral ibuprofen: a comparison of efficacy in the treatment of acute lower back injury. *Br J Clin Res* 1993;4:5—17.
63. Hurri H. The Swedish back school in chronic low back pain: part I: benefits. *Scand J Rehabil Med* 1989;21:33—40.
64. Meek JB, Giudice VW, Enrick NL. Colchicine highly effective in disk disorders. *J Neur Orthop Med Surg* 1984;5:215—20.
65. Rasmussen GG. Manipulation of low back pain (a randomized clinical trial). *Man Med* 1979;1:8—10.
66. Turner JA. Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. *J Consult Clin Psychol* 1982;50:757—65.
67. Waterworth RF, Hunter IA. An open study of diflunisal, conservative and manipulative therapy in the management of acute mechanical low back pain. *NZ Med J* 1985;98:372—5.
68. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. *Spine* 1993; 18:1433—8.
69. Wiesel SW, Cuckler JM, Deluca F, Jones F, Zeide MS, Rothman RH. Acute low-back pain: an objective analysis of conservative therapy. *Spine* 1980;5:324—30.
70. Wilkinson MJB. Does 48 hours' bed rest influence the outcome of acute low back pain? *Br J Gen Pract* 1995;45:481—4.
71. Wreje U, Nordgren B, Åberg H. Treatment of pelvic joint dysfunction in primary care — a controlled study. *Scand J Prim Health Care* 1992;10:310—5.

72. Mulder PH. Exercise therapy and musculoskeletal disorders: rheumatoid arthritis, osteoarthritis and back pain. Utrecht: NIVEL, 1992. No 42.
73. Aronoff GM, McAlary PW, Witkower A, Berdell MS. Pain treatment programs: do they return workers to the workplace? [review]. In: Hanley JJ, editor. Occupational Medicine: state of the art reviews, vol 3, no 1. Philadelphia (PA): Hanley & Belfus Inc, 1988:123—36.
74. Cutler RB, Fishbain DA, Rosomoff HL, AbdelMoty E, Khalil TM, Rosomoff RS. Does nonsurgical pain center treatment of chronic pain return patients to work?: a review and meta-analysis of the literature. *Spine* 1994;19:643—52.
75. Flor H, Fydrich T, Turk DC. Efficacy of multidisciplinary pain treatment centers: a meta-analytic review. *Pain* 1992;49: 221—30.
76. Lechner DE. Work hardening and work conditioning interventions: do they affect disability? [review]. *Phys Ther* 1994; 74:471—93.
77. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385—9.
78. Cats-Baril WL, Frymoyer JW. Identifying patients at risk of becoming disabled because of low-back pain: the Vermont Rehabilitation Engineering Center predictive model. *Spine* 1991;16:605—7.
79. Troup JDG. Causes, prediction and prevention of back pain at work. *Scand J Work Environ Health* 1984;10:419—28.
80. Loisel P, Durand P, Abenhaim L, Gosselin L, Simard R, Turcotte J, et al. Management of occupational back pain: the Sherbrooke model: results of a pilot and feasibility study. *Occup Environ Med* 1994;51:597—602.
81. Scheer SJ, Radack KL, O'Brien DR. Randomized controlled trials in industrial low back pain relating to return to work: part 1: acute interventions. *Arch Phys Med Rehabil* 1995;76: 966—73.

Appendix

Guidelines for assessment

1. Randomization

Appropriate:

- if the study participants have the same chance of receiving each intervention and the investigators cannot predict which treatment is next.

Not appropriate:

- date of birth
- date of admission
- hospital numbers
- alternation
- method not described

2. Blinding of study participant

Appropriate:

- if the study participant cannot identify the specific intervention
- if the treatment is presented as equally effective
- if, in the absence of equally effective treatment, the use of active placebos, identical placebos, or dummies is mentioned

3. Comparability or adjustment for incomparability

Appropriate:

- if the number of workers on sick leave (or the outcome parameter for vocational status in the study) and 3 or

more of the following 6 factors have been investigated in the study initially: number of workers with workers' compensation, duration of complaints or sick leave, age, occupation or physical work load, radiation or nerve root compression, previous episodes or recurrences. These factors should not differ in favor of the experimental group; when there is a significant difference, a sensitivity analysis or adjustment must be carried out.

- if a separate analysis of workers on sick leave has been carried out.

4. Loss to follow-up

Appropriate:

- when > 80% of the randomized patients minus correctly excluded patients after randomization has been included in the analyses.

5. Analysis for intention to treat

Appropriate:

- when loss to follow-up is <20% and all the randomized patients minus missing values patients, have been analyzed irrespective of noncompliance and cointerventions. When loss to follow-up is ≥ 20%, an alternative analysis which accounts for missing values (eg, worst case analysis) has to be incorporated.

Received for publication: 29 July 1996